

INFRARED CHEMICAL IMAGING FOR PATHOLOGY AND FORENSIC BIOLOGY

FIONA BURGER

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR
OF PHILOSOPHY (SCIENCE)

2013

DEPARTMENT OF CHEMISTRY AND FORENSIC SCIENCE

UNIVERSITY OF TECHNOLOGY, SYDNEY

CERTIFICATE OF AUTHORSHIP / ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Fiona Burger

2013

ACKNOWLEDGEMENTS

As with any large undertaking during one's life, the final product represents the work and dedication of not just the author, but the essential body of souls that each contribute, either directly or indirectly, to the research process as a whole.

Firstly to my supervisors and project partners thanks to whom which none of this work would ever have been possible without their valued time and input. To Dr. Brian Reedy, my primary supervisor over these many long years, thank you for sharing with me with your guidance and extensive pool of knowledge. Without your patience and perseverance with my erratic attendances I would still be stumbling in the dark. Thank you also to Dr. Philip Doble, my co-supervisor. To Dr. Richard Scolyer and Professor John Thompson of the Royal Prince Alfred Hospital Melanoma Research Unit, Mr. Kenneth Lai (formerly with the RPA Melanoma Research Unit), Dr. Neil Langlois (formerly with the Institute of Clinical and Medical Pathology Research - Westmead), Dr. Tamara Stzynda and Dr. Sarah Tarran, thank you for allowing me the time, samples and nevermind the whole impetus for these studies, as well as for always providing me with answers to my seemingly never-ending naïve pathology questions.

I must also mention the invaluable help that both Helen Rutledge and Dominic Hare, my fellow postgraduate students at UTS, provided to me during the course of my research work. Helen, your ability to make the often daunting and incomprehensible nature of chemometric work achievable and logical to me may only be described as a phenomenal skill. Your ability to understand my data handling requirements and adapt numerous programs and statistical methods to my requirements was invaluable, as was your patience and friendliness throughout. Thank you also Dominic; not only for the assistance with my sample gathering and preparation, but for also giving me that gentle push (whether you intended it or not) to get my work moving along.

And finally to my family and friends I must deliver a huge thanks. To Christie Wallace-Kunkel and Garry Sarkissian, my fellow PhD peers; you have shown to me that the theory of a completing a PhD part-time is possible, no matter the work or personal interruptions that may fall in our path and that with perseverance anything is achievable. To Paul, my long-suffering partner, thank you for never letting me give up and always reminding me of the greater goals in life. Thank you to my parents who, like Paul, have had to put up with the good, the bad, and the ugly from me for so many years now, and who never failed to offer encouragement, love and support when I needed it the most. Also, a special thanks to my mother for allowing me to join her on her own melanoma journey.

PREFACE

Tissue pathology assessment is primarily performed using simple human visual examination and interpretation. While various physico-chemical or biochemical treatments may be used to enhance or highlight components within the sample, the ultimate conclusion regarding the 'identity' of the sample lies with the very human examiner making comparisons with samples of known origin. Although this process sounds straightforward, more often than not the samples are less than ideal, and will be subject to natural variations which can introduce a whole gamete of artefacts which the examiner must integrate into their interpretation process. Suddenly, the comparison process goes from distinct, 'black and white' decisions, to an entire greyscale in-between.

Pathologists garner their knowledge not just from theory, but by training themselves over many years to recognise the normal from the abnormal. The pathologist must be able to deliver, with confidence, a diagnosis or ruling to a patient, treating medical professional, or (in some cases) a court. Their findings are based on a comparison of the subject tissue with many tissue samples of known prognosis, and the recognition of cell and tissue structures that 'do not belong'. Hence, it is years of exposure to endless numbers of cases which proves to be critical in achieving consistent and correct diagnoses.

Any tool that can improve the visual appearance of a sample in such comparison work is always of interest to the practitioner. Reducing the subjective nature of any scientific examination process, and more importantly reducing the potential for human error, is something all researchers strive for. The ability to confirm a pathological finding with an objective, computerised result, may also offer greater confidence to the information receiver, be they a patient or jury. It is the development of a more objective approach to tissue analysis that forms the fundamental premise of this work.

TABLE OF CONTENTS

CERTIFICATE OF AUTHORSHIP / ORIGINALITY	i
ACKNOWLEDGEMENTS	ii
PREFACE.....	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	x
LIST OF TABLES	xxiii
ABBREVIATIONS	xxvi
ABSTRACT	xxxii
INTRODUCTION	1
 CHAPTER 1: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY	 8
1.1 PATHOLOGY.....	9
1.1.1 The Pathologist.....	10
1.1.2 Current Issues in the Pathology Discipline	10
1.1.3 Histology	11
1.1.3.1 <i>Mammalian Tissue (a general overview)</i>	12
1.1.3.2 <i>Tissue Preparation</i>	12
1.1.3.3 <i>Preparation Artefacts</i>	17
1.1.3.4 <i>Tissue Cutting</i>	18
1.1.3.5 <i>Tissue Staining</i>	19
1.1.4 Histochemistry.....	20
1.1.5 Immunohistochemistry	21
1.1.6 The “Gold Standard” of Pathology	22
1.1.7 Histology - Where do we go from here?.....	23
1.2 INFRARED SPECTROSCOPY AND CHEMICAL IMAGING.....	25
1.2.1 Electromagnetic Radiation.....	25
1.2.2 Visible Light	27
1.2.3 Infrared Light.....	28
1.2.4 Infrared Instrumentation and Measurement Theory	31
1.2.5 FTIR Microspectroscopy	33
1.2.6 Applications of FTIR to Biological Studies	36

1.2.7	FTIR Chemical Imaging.....	41
1.2.8	Limitations of FTIR Chemical Imaging.....	44
1.2.9	Advantages of FTIR Chemical Imaging.....	45
1.2.10	Data Handling Methods.....	47
1.2.11	Data Preprocessing.....	49
1.2.11.1	<i>Baseline Correction</i>	49
1.2.11.2	<i>Spectral Smoothing</i>	49
1.2.11.3	<i>Derivatives</i>	51
1.2.11.4	<i>Normalisation</i>	53
1.2.11.5	<i>Spectral Subtraction</i>	54
1.2.11.6	<i>Mean Centring</i>	55
1.2.11.7	<i>Minimum Noise Fraction</i>	56
1.2.11.8	<i>Additional Preprocessing Steps</i>	56
1.2.12	Classification and Pattern Recognition.....	57
1.2.13	Common Multivariate Analysis and Pattern Recognition Methods.....	59
1.2.13.1	<i>Principal Component Analysis (PCA)</i>	59
1.2.13.2	<i>Cluster Analysis</i>	65
1.2.13.3	<i>Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)</i>	78
1.2.13.4	<i>Soft Independent Modelling of Class Analogy (SIMCA)</i>	80
1.2.13.5	<i>Artificial Neural Networks (ANN)</i>	85
1.2.13.6	<i>Vertex Component Analysis (VCA)</i>	90
1.2.14	Data Reduction and Feature Extraction.....	92
1.2.15	Test Sets and Validation of the Pattern Recognition Model.....	94
1.2.16	Selection of the Pattern Recognition Method.....	98
CHAPTER 1: REFERENCES.....		100
CHAPTER 2: MELANOMA STUDY PART 1 (<i>BACKGROUND AND METHOD</i>).....		117
2.1	MELANOMA THEORY.....	118
2.1.1	Normal Cell Growth.....	118
2.1.2	Disorders of Cell Growth.....	119
2.1.3	Cancer.....	121
2.1.4	Skin Cancer.....	123
2.1.5	Melanoma Epidemiology.....	124

2.1.6	Melanoma Pathogenesis	127
2.1.7	Melanoma Aetiology	131
2.1.8	Melanoma Diagnosis and Prognosis.....	135
2.1.9	Treatment.....	148
2.1.10	Accuracy and Reliability of Diagnosis	153
2.1.11	Melanoma Imaging	156
2.1.12	Data Analysis and Classification Automation.....	165
2.1.13	Project Research Goals.....	169
2.2	EXPERIMENTAL METHOD.....	171
2.2.1	Ethics Approval.....	171
2.2.2	Sample Selection and Preparation.....	171
2.2.3	Instrument and Image Collection Parameters.....	173
2.2.4	Data Processing	174
CHAPTER 2: REFERENCES.....		179
 CHAPTER 3: MELANOMA STUDY PART 2 (<i>RESULTS AND DISCUSSION</i>).....		196
3.1	RESULTS	197
3.1.1	Tissue Sections - Visible Features.....	197
3.1.2	Image Collection	198
3.1.3	Data Preprocessing	205
3.1.4	Spectral Reference Library Construction	206
3.1.5	Spectral Similarity Investigation.....	213
3.1.6	Classification Results.....	224
3.2	DISCUSSION.....	239
3.2.1	Reference Spectra Analysis and Comparison	239
3.2.2	Spectral Similarity Investigation.....	243
3.2.3	Image Classification.....	245
3.2.3.1	<i>Library Construction and Image Classification Observations.....</i>	<i>245</i>
3.2.3.2	<i>Impact of Paraffin on Image Classification.....</i>	<i>248</i>
3.2.3.3	<i>Method Challenges and Future Directions.....</i>	<i>249</i>
3.3	CONCLUSION	254
CHAPTER 3: REFERENCES.....		256

CHAPTER 4: WOUND STUDY PART 1 (BACKGROUND AND METHOD).....	261
4.1 WOUND THEORY.....	262
4.1.1 Definition of a Wound.....	262
4.1.2 The Importance of Wound Examination and Age Determination.....	266
4.1.3 The Wound Ageing Process.....	269
4.1.4 Skin - Function and Structure.....	270
4.1.4.1 <i>Function</i>	270
4.1.4.2 <i>Structure</i>	271
4.1.5 Wound Healing.....	274
4.1.6 Types of Wound Repair.....	286
4.1.7 Previous Wound Ageing Studies.....	288
4.1.8 Practical Considerations in Wound Ageing Studies.....	297
4.1.9 Infrared Spectroscopy and Wound Ageing.....	300
4.1.10 Wound Type Selection.....	307
4.1.11 Animal Subject Selection.....	307
4.1.12 Sample Size.....	311
4.2 EXPERIMENTAL METHOD.....	312
4.2.1 Ethics Approval.....	312
4.2.2 Animal Model.....	312
4.2.3 FTIR Study Sample Preparation.....	317
4.2.4 Instrument Parameters and Image Collection.....	317
4.2.5 Data Pre-processing and Analysis.....	319
CHAPTER 4: REFERENCES.....	321
 CHAPTER 5: WOUND STUDY PART 2 (RESULTS AND DISCUSSION).....	 333
5.1 PILOT STUDY.....	334
5.1.1 Results.....	334
5.1.1.1 <i>Skin Tissue Sections - Visible Features</i>	334
5.1.1.2 <i>Skin Tissue Sections - Infrared Features</i>	334
5.1.1.3 <i>Data Preprocessing Approaches</i>	344
5.1.1.4 <i>Principal Component Analysis / Linear Discriminant Analysis (PCA-LDA)</i>	344
5.1.1.5 <i>Soft Independent Modeling of Class Analogy (SIMCA)</i>	350
5.1.2 Discussion.....	350

5.1.2.1	<i>Principal Component Analysis / Linear Discriminant Analysis...</i>	353
5.1.2.2	<i>Soft Independent Modeling of Class Analogy</i>	356
5.1.2.3	<i>Pilot Study Result Summary.....</i>	357
5.2	MAIN STUDY	359
5.2.1	Results	359
5.2.1.1	<i>Data Preprocessing Approaches</i>	359
5.2.2	Discussion.....	361
5.2.2.1	<i>Principal Component Analysis / Linear Discriminant Analysis..</i>	361
5.2.2.2	<i>Soft Independent Modeling of Class Analogy</i>	368
5.2.2.3	<i>Self Organising Map (SOM) Investigation</i>	368
5.2.3	Discussion.....	378
5.2.3.1	<i>Principal Component Analysis / Linear Discriminant Analysis..</i>	378
5.2.3.2	<i>Soft Independent Modeling of Class Analogy</i>	380
5.2.3.3	<i>Self Organising Map Algorithm.....</i>	383
5.3	MINIMUM NOISE FRACTION INVESTIGATION	386
5.4	GENERAL DISCUSSION.....	391
5.4.1	Comparison of Pilot Study to Main Study.....	391
5.4.2	Comparison of Infrared Study to Previous Wound Age Research	393
5.4.3	Further Considerations	397
5.5	CONCLUSIONS	398
	CHAPTER 5: REFERENCES.....	400
	CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS.....	402
6.1	CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH	403
6.1.1	Melanoma Study.....	403
6.1.2	Wound Study.....	406
6.1.3	Final Conclusions	409
	CHAPTER 6: REFERENCES.....	411

LIST OF FIGURES

CHAPTER 1: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY.....	8
1-1 Formalin fixation results in the formation of a stabilising meshwork caused by the cross-linking of proteins. The reaction above depicts the formation of a methylene bridge during the fixation of a tissue with formalin	13
1-2 (a) Example of a plastic histology tissue cassette; (b) a paraffin-embedded tissue block (front and back).....	15
1-3 Schemes of progression of paraffin sections from de-waxing xylene to staining and dehydration and clearing to mounting.....	16
1-4 Examples of unstained and H&E stained tissue. Image (a) demonstrates a lack of contrast which makes cell and tissue type identification and observation difficult. Image (b) shows the improvement gained from employing a simple tissue stain such as H&E.....	20
1-5 Example of the bad publicity and treatment consequences from erroneous results	23
1-6 Diagrammatic representation of the dual nature of the electromagnetic wave. Note that the electric and magnetic fields are perpendicular to each other as well as both being transverse to the direction of the wave propagation.....	25
1-7 The electromagnetic spectrum. The boundaries between the bands depicted are arbitrary and do not delineate strict borders	26
1-8 The three main regions of the infrared portion of the electromagnetic spectrum and their approximate corresponding wavenumber ranges.....	29
1-9 Michelson interferometer	32
1-10 Diagrammatic representation of the construction of a FTIR microscope	34
1-11 Diagrammatic representation of the ‘transflection’ sample analysis method (from Bassan <i>et al</i>).....	38
1-12 Diagrammatic representation of the three dimensional image data array (sometimes referred to as a “data cube”).....	39
1-13 (a) Schematic of an indium antimonide (InSb) detector heterostructure, and (b) a complete focal array detector assembly.....	42

1-14	An example of the spectral complexity of a mammalian tissue sample - coronal mouse brain section. Panel B is a coronal section of brain imaged at 1650 cm^{-1} (amide I peak), 16 cm^{-1} resolution in reflectance mode. Panels C and D show the spectra obtained from two regions of the tissue sample.....	48
1-15	An example of how the data points are selected for a 5-point moving average	50
1-16	Spectral derivatives: (a) a single absorption peak in the original spectrum; (b) the first derivative and (c) the second derivative.....	51
1-17	Example of how overlapping peaks may be resolved taking the second derivative (and multiplied by -1) of the original spectrum.....	52
1-18	(a) Basic two-dimensional Cartesian coordinate representation of measurement (x_i , y_i). (b) The same measurement as in (a), but represented as a data vector, \vec{t} (adapted from Miller and Miller).....	58
1-19	Simple two-dimensional representation of the determination of principal components.....	60
1-20	A diagrammatic explanation of Equation 9 (adapted from Brereton)	62
1-21	Examples of an eigenvalue (Graph A) and score plot (Graph B) obtained using Principal Components Analysis with a covariance matrix.....	63
1-22	An example of a data set where PCA does not give group separation of data	65
1-23	Two vectors t and r representing two spectra. α describes the angle between the vectors.....	67
1-24	Two-dimensional graphical representation of the Euclidean distance measure between two vectors.....	69
1-25	A simplistic diagram of the clustering process adapted from Miller and Miller. Figure (a) depicts the original data, and Figures (b) to (e) show the sequential clustering steps performed.....	73
1-26	A dendrogram for the clustering performed in Figure 1-25. The table to the right of the dendrogram lists the corresponding amalgamation step data from Minitab using the Euclidean distance measure and single linkage.....	75

1-27	Diagrammatic depiction of the k-means clustering work flow (adapted from Brereton).....	76
1-28	A diagrammatic representation of how LDA may be used to distinguish between two classes (taken from Brown).....	79
1-29	Example of how two models obtained using SIMCA may use a different number of principal components to describe a given group of data. The group k data is modeled using one principal component, whereas group l uses two principal components.....	81
1-30	Example of a Cooman's plot with data groups k and l	85
1-31	An example of a feed forward neural network. The flow of data through the network structure is depicted with arrows to show the dependencies between variables.....	86
1-32	Graphs A and B are examples of data for which neural networks can provide complex decision boundaries. A distinct advantage with ANNs is their ability to recognise patterns in data that more conventional statistical algorithms may fail to perceive.	88
1-33	SOM application example adapted from Köhler <i>et al.</i> In this example a simple three-dimensional data set (a) has been analysed (the red symbols represent data points, the black prototype vectors) using SOM to produce (b) a U-matrix. The colour scale indicates the prototype vector distances of directly connected units. Accordingly, cluster borders are indicated by the presence of warmer colours (larger mean distances between vectors). In this diagram the three data clusters are clearly visible.....	90
1-34	VCA algorithm diagram depicting the vector space for three wavenumber values. Note that all of the vector endpoints lie in a triangle. Adapted from Matthäus <i>et al.</i>	91
1-35	Three possible models for a data set (with non-overlapping groups) are represented by the three circles, with the locations of the training and test samples indicated. While the innermost circle has the lowest standard error, some of the reference samples from that group are not included. The outermost circle is too wide a model as it includes samples from a different group (adapted from Flåten).....	94

1-36	Example of a data analysis process workflow which employs LDA.....	95
CHAPTER 2: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY.....		117
2-1	A diagrammatic representation of the cell cycle.....	118
2-2	Examples of the abnormalities of cell growth and maturation that may be observed in tissues (adapted from Chandramosa and Taylor).....	120
2-3	A recent revision of the hallmarks of cancer. Traditionally there were six recognised hallmarks of cancer, but some researchers argue that inflammatory cells and molecules are important in influencing cancer progress and should also be included (from Mantovani).....	121
2-4	Images of various skin cancers (A) melanoma, (B) nodular melanoma, (C) squamous cell carcinoma and (D) basal cell carcinoma.....	124
2-5	A graph showing the international variation of melanoma incidence from Tracey <i>et al.</i>	126
2-6	The most common types of cancer in the Australian NSW population in 2007 from Tracey <i>et al.</i>	126
2-7	A diagrammatic comparison of normal skin (s) to a melanocytic naevus lesion possessing a dysplastic component (morphologically atypical melanocytes) (b). From Gray-Schopfer <i>et al.</i>	127
2-8	The phases of histologic progression of melanocyte transformation using a pictorial (left) and a corresponding H&E-stained section (right) at magnification x20 (scale bar: 20µm).	129
2-9	Diagrammatic representation of the divergent pathway model for cutaneous melanoma from Whiteman <i>et al.</i>	130
2-10	The lymphatic system (from Clark and Grant).....	132
2-11	A diagrammatic representation of the relationship between the risk of skin cancer (melanoma, BCC and SCC) and ultraviolet radiation exposure (from Gordon and Hirst).....	135
2-12	Examples of haematoxylin-eosin staining (A) and MART-1 immunostaining (B) on a section of tissue exhibiting superficial spreading melanoma. Original magnification x100. Image from Forman	140
2-13	Recommended method for sampling a lymph node for histological analysis	

	(the numbers on the right indicate the required serial full face sections). Note however that larger nodes (> 10 mm thick) may require additional parallel slices to be cut.....	142
2-14	The Clark levels of invasion (from Cochran and Crotty).....	147
2-15	Flow chart representing the approach taken for the infrared imaging of tissues for melanoma diagnosis (adapted from Bird <i>et al</i>).....	178
CHAPTER 3: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY.....		196
3-1	An example of an H&E-stained serial section of a lymph node provided by the Sydney Melanoma Unit. The black markings were made by a pathologist on the side to indicate the tumour and 'normal' lymph node (LN) regions.	197
3-2	(A) An example H&E stained section of naevus human skin tissue at x1.6 objective magnification. (B) A magnified region of the naevus tissue displaying the dense accumulation of pigmented melanin cells at x4.0 objective magnification. The epidermis E, dermis D and the naevus growth N are all clearly visible. (C) A closer view of the naevus growth displaying the dense accumulation of pigmented melanin cells at x10 objective magnification.....	199
3-3	(A) An example H&E stained section of a human skin tissue section exhibiting primary melanoma at x5 objective magnification. (B) A closer view of a region of the same tissue sample displaying melanoma cell growth. (C) A magnified (x10 objective) region of the same tissue sample. Note the nests of atypical melanocytes N occupying all levels of the epidermis which are considered typical of superficial spreading melanoma.....	200
3-4	(A) An example of a normal lymph node from Fiore. (B) A x5 objective magnification of an H&E stained section of a normal human lymph node section used in this study. The capsule C, cortex X and the medulla M are all clearly visible.....	201
3-5	(A) A human lymph node at x0.63 objective magnification, but in this case the sample is exhibiting metastatic involvement. (B) Magnified region of the metastatic lymph node (x1.0 objective magnification) from (A) shows distinct differences from the normal lymph node tissue section observed in Figure 3-	

4.	The marked rectangles correspond with further magnification of the selected tissue regions displayed in Figure 3-6.....	202
3-6	Magnified regions (x5 objective magnification) (a) and (b) marked in Figure 3-5 showing the variety of cell types and shapes observed in the metastatic lymph node. (c) & (d) are x20 objective magnification images of regions (a) and (b) respectively. The cells in (d) are not clearly distinct entities and in particular exhibit irregular shapes and lack a clearly staining nucleus, particularly when contrasted to the tissue shown in (c).....	203
3-7	An example of the 'tiling effect' observed in a lymph node sample and the degree to which its appearance varies when the image is viewed at different wavenumber values. This effect may be even more visible following image classification if effective data preprocessing is not employed. The source of the effect is attributed to a number of causes which include subtle stage calibration issues, uneven illumination of the sample, as well as background collection limitations (using a 4-tile background for a 16-tile image) and instrumental variations over the long sample collection times	207
3-8	(A) An example of one of the training samples used in the spectral reference library construction - an H&E-stained section of a lymph node exhibiting metastatic melanoma. (B) The HCA clustering image showing the first 13 clusters, some of which were then selected for inclusion in the first library construction method. (C) Examples of the ROI selection method conducted in CytoSpec for the collection of average spectra for the second library construction method explored. The average spectra collected during this stage have been presented in Figure 3-9	209
3-9	Comparison of the average second derivative spectra (offset from each other to assist in ease of viewing) for the three "tissue types" sampled using the ROI regions marked in Figure 3-8C. Two ROIs (i.e. non-contiguous regions) were obtained for each of the three tissue types based on a comparison with the H&E-stained slides in this case rather than relying solely on the HCA cluster regions.....	210
3-10	Example of the training images obtained using VCA from a deparaffinised lymph node tissue section exhibiting metastatic melanoma growth (pre-processed data block).....	211

3-11	Average spectra (offset for convenience of display by 0.5 absorbance units) from the different tissue groups targeted in this study (spectral reference library construction Method 1). Note the high similarity in the appearance of the spectra between the various tissue regions sampled.....	214
3-12	Multiplot of class distance for reference library Method 2 (ROI) utilising deparaffinised tissue and the second derivative (1012 - 1585 cm^{-1} spectral region). Samples in the top left and bottom right regions indicate clear membership to the x and y axis classes. Samples in the bottom left region indicate samples which could belong to either class and those in the top right region belonging to neither	218
3-13	Total Modelling Power plot for reference library Method 2 (ROI) utilising deparaffinised tissue and the second derivative (1012 - 1585 cm^{-1} spectral region). It is clear that there is no one specific region of the spectrum which is of greatest importance for any class in the spectral data set examined	219
3-14	Total Discriminating Power plot for reference library Method 2 (ROI) utilising deparaffinised tissue and the second derivative (1012 - 1585 cm^{-1} spectral region).....	219
3-15	PCA/LDA discriminant score function plot for reference library Method 1 (HCA) utilising paraffin-embedded tissue and the original spectra using the 1000 - 1585 cm^{-1} spectral region. The degree of overlap between the training sample spectra is considerable and strongly indicates a high degree of similarity between the tissue spectra.....	220
3-16	Multiplot of class distance for reference library Method 2 (ROI) utilising deparaffinised tissue and the second derivative (1012 - 3086 cm^{-1} spectral region). The tissue spectra have been sub-classified within the four main training tissue types of lymph nodes exhibiting metastatic melanoma (M_L), normal lymph nodes (N_L), skin tissue containing primary melanoma (M_S) and skin tissue with a naevus component (N_S)	223
3-17	PCA/LDA discriminant score function plot for reference library Method 2 (ROI) utilising deparaffinised tissue and the second derivative spectra using the 1012 - 3086 cm^{-1} spectral region. Although there is overlap between the different tissue types (i.e. skin and lymph node) there are distinct groupings	

	of tissue from the same sub-classification type (e.g. epidermis spectra from both skin sample types).....	224
3-18	Demonstration of the SAM results using the various spectral reference libraries for a training image (lymph node tissue section exhibiting metastatic melanoma growth, deparaffinised tissue, second derivative data block). The left hand images use all library reference spectra, irrespective of tissue type source. The right hand images use spectra extracted only from lymph node training images.....	226
3-19	Demonstration of the SAM results using the various spectral reference libraries for a test image (normal lymph node, deparaffinised tissue, second derivative data block). Left images = all tissue library reference spectra; right images = lymph tissue library spectra only	227
3-20	Classification of a naevus skin tissue section (deparaffinised, 2 nd derivative data block) using reference library construction Method 1 and SAM. The presence of primary melanoma identified (and hence misclassified) pixels in a known naevus sample is a matter for concern and demonstrates clearly that the classification approach is imperfect but the majority of pixels have at least been correctly classified in this example.....	229
3-21	An example of a test skin tissue with primary melanoma present sample image exhibiting good quality classification using library construction Method 1 and SAM. Unfortunately the 'tiling effect' may also be observed to varying extents in the classified images	231
3-22	An example of a poorly classified test sample (skin tissue section exhibiting naevus growth) using the library construction Method 1 and SAM. A large number of pixels have been misclassified as primary melanoma (yellow)	232
3-23	Classification of a primary melanoma skin tissue section (pre-processed data block) using library construction Method 2 and SAM. The use of the spectral library containing all tissue spectra clearly does not perform as well as the tissue-specific library, although the poor classification observed in the 'all' library for the training image classification was not so pronounced in the other samples examined.....	233

3-24	SAM classification using library construction Method 2 of a lymph node sample exhibiting melanoma metastasis (pre-processed data block). Red pixels indicate spectra classified as metastatic melanoma. Cortex spectra sampled from lymph nodes exhibiting metastatic melanoma involvement are indicated by pixels sea green in colour. Truncation of the image spectra to exclude the amide I peak resulted in overall classification improvement in a large number of the samples examined using this library method.....	234
3-25	An example where the original data block produced a better image classification than that of the corresponding derivative data block. This result was generally uncommon and observed only in paraffin-embedded tissue images	235
3-26	A naevus tissue sample processed using library construction Method 3 and the SAM classification algorithm. Purple pixels indicate classification as naevus tissue, yellow as primary melanoma. Note the generally poor overall tissue classification for both data blocks and in particular the original data.....	238
3-27	Original (i.e. unprocessed) average spectra of the skin (left plot) and lymph node (right plot) spectra obtained from the training tissue sample images. The spectra have not been subjected to any form of smoothing or normalisation at this stage and is in its most raw format.....	240
3-28	A comparison of spectra taken from the approximately the same region of a lymph node prior to and following deparaffinisation. The paraffin spectrum was obtained from the tissue block residue present on a tissue section slide prior to deparaffinisation. The large bands attributable to paraffin are an obvious problem considering where some of the spectral bands of interest are located in biological tissues	248
CHAPTER 4: WOUND STUDY PART 1 (BACKGROUND AND METHOD).....		261
4-1	Some examples of pointed and sharp edged weapons and the appearance of the skin wounds they produce (adapted from Di Maio)	263
4-2	A diagram showing the time course from injury to autopsy.....	269
4-3	An example of an H&E-stained section of skin showing the epidermis and dermis	272
4-4	The multiple layers of the skin.....	272

4-5	(a) Diagrammatic representation of the arrangement of the dermal papillae, and (b) an H&E-stained section of skin showing the papillary and reticular layers.....	273
4-6	The three phases of wound healing and the order in which they occur. Note the overlap, particularly between that of the first two phases.....	275
4-7	A skin wound 24 - 48 hours after injury demonstrating the cell groups typically present during the inflammation stage of healing. A fibrin clot has filled the wound area and the wound environment is ischaemic.....	279
4-8	A skin wound undergoing the proliferation phase of wound healing. In this diagram the skin is shown 5 - 10 days after injury showing the eschar (scab) formed over the wound surface. New blood vessels have formed in the wound region and the basal cells of the epidermis are migrating across the wound gap.....	282
4-9	A skin wound after 1 - 12 months of healing (undergoing the remodeling phase). Note that the wound has contracted near its surface and the new epidermal layer appears slightly higher than that of the surrounding surface. Within the dermal region of the wound collagen (still disorganised in configuration) has been deposited by fibroblasts, but there are no normal skin appendages present.....	283
4-10	Timeline showing the various wound healing phases and some of the major cellular groups involved, taken from Lorenz and Longaker.....	284
4-11	A flow chart showing normal wound healing, adapted from Mandracchia <i>et al</i>	285
4-12	Healing by Primary (First) and Secondary (Second) Intention.....	287
4-13	Schematic diagram showing the zones of an antemortem injury.....	289
4-14	Timing of injuries	291
4-15	Diagrammatic representation of the rat wounding sites (adapted from Tarran)..	315
4-16	Flow diagram of Animal Model procedures (obtained from Tarran).....	316
4-17	Region of wound tissue targeted for imaging.....	318
CHAPTER 5: WOUND STUDY PART 2 (RESULTS AND DISCUSSION).....		333
5-1	(A) An example H&E stained section of unwounded (control) rat skin at x5 objective magnification. The epidermis (a), dermis (b), subcutaneous tissue	

	(c) and the upper edge of the panniculus carnosus (muscle) (d) layers are all clearly visible. (B) 5 minute wound at x5 objective magnification. (C) 30 minute wound at x5 objective magnification. (D) A magnified (x20 objective) section of the 30 minute wound.....	335
5-2	(A) 3 hour wound. (B) 6 hour wound. (C) 12 hour wound and an enlargement (D) displaying the large concentration of blood cells still present in the wound fissure	336
5-3	(A) 24 hour wound and enlargement of wound region in (B). (C) 120 hour (5 day) old wound. The wound appears to be well opposed by this stage..	337
5-4	(A) 168 hour (7 day) old wound. Wound closure and re-epithelialisation has occurred. (B) A x20 objective magnification of the boxed section in B which clearly shows a portion of the thickened epidermis and the region of healed dermis below. As with the 5 day old wound, the actual wound has become difficult to detect. (C) 7 day old wound. The wound site (marked W) is difficult to detect due to the well-advanced healing that has taken place	338
5-5	(a) Large mosaic image of a 24 hour antemortem wound tissue section as viewed using FTIR CI (8 cm^{-1} spectral resolution, displayed at 1543 cm^{-1}), and (b) example infrared spectra sampled from various regions of the image given in (a). The spectra have been off-set from each other for ease of viewing	339
5-6	An enlarged view of the spectral region $900 - 1700\text{ cm}^{-1}$ for the spectra shown in Figure 5-5b. The muscle, epidermal and dermal layers all possess strong protein component peaks.....	340
5-7	(a) A large mosaic image of a 5 day wound (8 cm^{-1} spectral resolution, image displayed at 1547 cm^{-1}). The spectra for markers A and B are given in (b). The high protein contribution and reduced noise observable in marker A's spectrum is consistent with the expected elevated concentration of protein due to the dense packing of collagen fibres formed in the wound gap following wound edge apposition. Marker B is from a region of dermal tissue in the area adjacent to the main wound site and is more typical of the spectra observed in 'normal' dermis with its 'loose' connective tissue structure	343
5-8	Imaging and data reduction process overview employed in the Pilot Study.....	345

5-9	Linear discriminant analysis scores plot of the first and second score values for the wound data obtained using Method 1 (second derivative with 7 smoothing points followed by the MIQR feature vector being extracted). Ten principal component scores were input into SPSS for the subsequent LDA calculations	348
5-10	An example of a Scree plot and its associated eigenvalues produced during PC analysis in Minitab for wound data pre-processed using Method 16. For this data set, it would appear that the majority of the variance within the data is explained within the first few PCs.....	354
5-11	Example of the spectral quality test thickness test as implemented in CytoSpec using a 7 day old wound tissue section (16 cm^{-1} spectral resolution, imaged at $1620\text{-}1680\text{ cm}^{-1}$).....	360
5-12	(a) Example spectra obtained from a 2 day-old wound tissue section at the same location with varying spectral resolution values. The spectra have been off-set for display purposes. (b) Example of the effect on spectral quality when PCA-based noise reduction is performed on spectra of varying resolution values from (a). The first five principal components have been retained to produce these spectra. It is clear that this noise reduction approach has the greatest impact on the 4 cm^{-1} spectral resolution data.....	362
5-13	Imaging and data reduction process overview employed in the Main Study.....	363
5-14	Linear discriminant analysis scores plot obtained for the 8 cm^{-1} wavenumber spectral resolution ('average' feature vector) Main Study wound data which has undergone preprocessing according to Method 23. Fifteen principal component scores were provided for the subsequent LDA classification.....	371
5-15	SIMCA total modeling power for data preprocessing Method 26 (4 cm^{-1} resolution, 5 PCs noise reduction) for the training data set. It would appear that the median component of the feature vector contains more discriminatory features than the interquartile range region.....	381
5-16	SOM results for the 16 cm^{-1} resolution data using the Method 32 pre-processing routine and the MIQR feature vector. (a) The resulting U-matrix. (b) This diagram shows the number of data points (using the red hexagon's size to indicate frequency) associated with each SOM (best matching unit - BMU). c) Each SOM unit's associated data label.....	384

5-17	Supervised SOM results for the 4 cm ⁻¹ resolution wound data pre-processed using Method 33 (MIQR feature vector). In this case the MQE value did not change dramatically as it was already quite high.....	385
5-18	MNF comparison of the first three components for the Pilot Study training data (both original and pre-processed). (a) Depicts the layout of the data showing the five columns of pixels from each wound age group. Images (b) and (c) are the resulting RGB images constructed from the first three MNF-derived bands	388
5-19	Example of the scatterplots produced using the Pilot Study data during the MNF transform in ENVI and how they may be used to select endmembers manually. In (a), two ROIs have been selected from the 'arms' of the scatterplot cloud constructed from the first and third MNF bands. The spatial locations of these pixels are highlighted in the image in (b). (c) depicts the corresponding average spectrum for each ROI.....	389
5-20	Images (A), (B) and (C) are all sections taken from 24 hour wounds and clearly show how the rate of wound closure has not taken place consistently between the rat subjects	396
CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS.....		402
6-1	An example of an automated cancer diagnosis process using prostate tissue (from Kwak <i>et al</i>)	405

LIST OF TABLES

CHAPTER 2: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY	117
2-1 Recommended terminology for common forms of malignant melanoma	137
2-2 The AJCC system for melanoma of the skin	146
2-3 Key prognostic factors in melanoma	148
2-4 An example list of the data that may be included in a pathological report on a melanoma	149
2-5 Unlitigated and civil litigated claims - National Medical Defence Organisation (1990 - 2005)	154
2-6 An example of the large number of IR spectral features that are assigned to skin, obtained from Kong <i>et al.</i>	166
2-7 Tissue samples employed in this study	172
 CHAPTER 3: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY	 196
3-1 Summary of the spectral reference library construction approaches. ‘Original’ refers to the unprocessed data and ‘pre-processed’ refers the spectra subjected to the preprocessing described in Section 3.1.3.	212
3-2 SIMCA analysis of the library reference spectra grouped into the four source tissue types. First truncation refers to the removal of the C-H stretching region of the spectrum. Second truncation describes the subsequent removal of the Amide I band	216
3-3 PCA/LDA analysis of the library reference spectra grouped into the four sample type categories of lymph nodes exhibiting metastatic melanoma involvement, normal lymph node, primary melanoma and naevus tissue. The two values represent the percentage of original grouped cases correctly classified followed by the percentage of cross-validated grouped cases correctly classified in brackets	217
3-4 The twelve sub-classifications used for the more specifically labelled tissue regions within the training samples	221
3-5 SIMCA (% correctly classified) analysis of the library reference spectra	

	grouped into the twelve tissue sample type categories.....	222
3-6	PCA/LDA analysis of the library reference spectra grouped into the twelve tissue sample type categories. The two values represent the percentage of original grouped cases correctly classified followed by the percentage of cross-validated grouped cases correctly classified in brackets.....	222
3-7	Range of cluster values employed in the training sample images for library construction Method 1	228
3-8	Range of endmember values required for tissue region identification in the training sample images for reference spectra library construction Method 3.....	237
 CHAPTER 4: WOUND STUDY PART 1 (BACKGROUND AND METHOD).....		261
4-1	The principal growth factors, chemoattractants, and vasoactive substances involved in the wound healing process (other than those involved in clotting and complement cascades) (taken from Iocono <i>et al</i>)	276
4-2	Examples of potential wound age and vitality markers examined in the literature.....	298
4-3	Physiological comparison of rats to humans from Dorsett-Martin.....	310
4-4	Wound ages used in the Pilot Study	313
4-5	Wound ages used in the Main Study.....	314
4-6	Instrument operating parameters.....	318
 CHAPTER 5: PATHOLOGY AND INFRARED CHEMICAL IMAGING THEORY.....		333
5-1	Data preprocessing combinations applied to the Pilot Study wound data.....	347
5-2	Top five PCA-LDA classification results for the Pilot Study data based on test set classification accuracy. In each case, the 3 hour wound test set sample was incorrectly classified as a 1 hour old wound	348
5-3	Best PCA-LDA classification results achieved for each data pre-processing method trialed in the pilot wound study.....	349
5-4	Top five SIMCA classification results for the Pilot Study data based on test set and training model data set classification accuracy. The incorrectly classified test data samples are shown	351

5-5	SIMCA classification results achieved for each data pre-processing method trialed in the pilot wound study	352
5-6	Data preprocessing combinations applied to the Main Study wound data	365
5-7	Top ten PCA-LDA classification results for the Main Study data based on training set classification accuracy	366
5-8	Top ten PCA-LDA classification results for the Main Study data based on test set classification accuracy	367
5-9	Top ten PCA-LDA classification results for the Main Study data based on training set classification with ± 1 group accuracy	369
5-10	Top ten PCA-LDA classification results for the Main Study data based on test set classification with ± 1 group accuracy	370
5-11	Top ten SIMCA classification results for the Main Study data based on training model data set classification accuracy	372
5-12	Top ten SIMCA classification results for the Main Study data based on test data set classification accuracy	373
5-13	Top ten SIMCA classification results for the Main Study data based on training model data set classification accuracy allowing for a wider classification range	374
5-14	Top ten SIMCA classification results for the Main Study data based on test data set classification accuracy allowing for a wider classification range	375
5-15	An example of the training data SIMCA classification output obtained using the 8 cm^{-1} spectral resolution wound data pre-processed with Method 57 (noise reduction: 15 PCs retained, ASD feature vector)	376
5-16	An example of the test data SIMCA classification output obtained using the 8 cm^{-1} spectral resolution wound data pre-processed with Method 31 (noise reduction: 15 PCs retained, ASD feature vector)	377
5-17	Pathologist-assigned grading of wound age of the rat skin tissue sections reported in Tarran's study	394

ABBREVIATIONS

%CC	Percentage of Correctly Classified
AAS	Atomic Absorption Spectroscopy
AIHW	Australian Institute of Health and Welfare
AJCC	American Joint Committee on Cancer
ANN	Artificial Neural Networks
ALM	Acral Lentiginous Melanoma
ANOVA	Analysis of Variance
ASCII	American Standard Code for Information Interchange
ASD	Average-Standard Deviation
AUC	Area Under the Curve
Avg	Average
BA	Boric Acid
BCC	Basal Cell Carcinoma
BMU	Best-Matching Unit
BTEM	Band-Target Entropy Minimisation
CC%	Correct Classification Percentage
CCD	Charge Coupled Device
CDK	Cyclin/cyclin-dependent kinase
CI	Chemical Imaging
CLND	Completion Regional Lymph Node Dissection
CT	Computed Tomography
DNA	Deoxyribonucleic acid
ED	Euclidean Distance
EFOV	Expanded Field Of View
EGF	Epidermal Growth Factor

ELISA	Enzyme-linked Immunosorbent Assay
ELND	Elective Lymph Node Dissection
EMSC	Extended Multiplicative Signal Correction
FGF	Fibroblast Growth Factor
FIR	Far-infrared
FISH	Fluorescence <i>in situ</i> Hybridisation
FNB	Fine Needle aspiration Biopsy
FPA	Focal Plane Array
FRCPA	Fellow of the Royal College of Pathologists of Australasia
FTIR	Fourier Transform Infrared
FTIR CI	Fourier Transform Infrared Chemical Imaging
G0	Gap 0
G1	Gap 1
G2	Gap 2
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
H&E	Haematoxylin and Eosin
HCA	Hierarchical Cluster Analysis
HLA	Human Leukocyte Antigen
HPLC	High Performance Liquid Chromatography
ICP-MS	Inductively Coupled Plasma-Mass Spectrometry
IFN	Interferon
IGF	Insulin-like Growth Factor
IL	Interleukin
IP	Intraperitoneal
ISEL	<i>In Situ</i> End labelling
LA-ICP-MS	Laser Ablation Inductively Coupled Plasma Mass Spectroscopy
LDA	Linear Discriminant Analysis

LDH	Lactate Dehydrogenase
LN	Lymph Node
LOO	Leave One Out
LT	Leukotriene
LX	Lipoxin
MALDI-IMS	Matrix-assisted Laser Desorption Ionisation Imaging Mass Spectrometry
MCT	Mercury Cadmium Telluride
Med	Median
mf	Myofibroblasts
MF	Matched Filtering
Min/Max	Minimum / Maximum
MIQR	Median-Interquartile Range
MIR	Mid-infrared
MMP	Matrix Metalloproteinase
MNF	Minimum Noise Fraction
MQE	Mean Quantisation Error
MRDF	Maximal Multinomial Logistic Regression
mRNA	Messenger Ribonucleic Acid
MTMF	Mixture-Tuned Matched Filtering
MVN	Multivariate Normal
m/z	Mass to Charge Ratio
NADH	Nicotinamide Adenine Dinucleotide
NIR	Near-infrared
NMSC	Non-melanoma Skin Cancer
NN	Nearest Neighbour
NOR	Nucleolar Organiser Region
NR-Laser-SNMS	Non-resonant Laser Secondary Neutral Mass Spectrometry

NSW	New South Wales
OPD	Optical Path Difference
PAP	Papanicolaou
PAS	Periodic Acid-Schiff
PBS	Phosphate-buffered Saline
PC	Principal Component
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PDGF	Platelet-derived Growth Factor
PG	Prostaglandin
PPI	Pixel Purity Index
pTNM	Primary Tumour Node Metastasis
QDA	Quadratic Discriminant Analysis
RCPA	Royal College of Pathologists of Australasia
RGP	Radial Growth Phase
RNA	Ribonucleic Acid
ROI	Region Of Interest
RPAH	Royal Prince Alfred Hospital
RT-PCR	Reverse Transcriptase-polymerase Chain Reaction
SAM	Spectral Angle Measure
SCC	Squamous Cell Carcinoma
SCS	Spectral Correlation Similarity
SIMCA	Soft Independent Modelling of Class Analogy
SLNB	Sentinel Lymph Node Biopsy
SMLR	Sparse Multinomial Logistic Regression
SN	Sentinel Node (<i>aka</i> Sentinel Lymph Node)
SOM	Self Organising Map

SPF	Specific Pathogen Free
SSV	Spectral Similarity Value
SVM	Support Vector Machine
SVP	Spectral Vector Properties
TE	Topographic Error
TGF	Transforming Growth Factor
THSIS	Thermal and Hyperspectral Imaging System
TIL	Tumour Infiltrating Lymphocyte
TNF	Tumour Necrosis Factor
TNM	Tumour (T) / Lymph Nodes (N) / Metastasis (M)
TOF-SIMS	Time-of-flight Secondary Ion Mass Spectrometry
UTS	University of Technology, Sydney
UV	Ultraviolet
UV-Vis	Ultraviolet-Visible
VCA	Vertex Component Analysis
VGP	Vertical-Growth Phase
WLE	Wide Local Excision

ABSTRACT

The objective of this research was to explore the capability of Fourier Transform Infrared Chemical Imaging (FTIR CI) for two specific pathology applications: 1) the analysis of human tissues for the diagnosis of melanoma, and 2) incised skin wound age determination for the purpose of forensic investigation.

For the melanoma study, thin serial sections were obtained from an archival tissue bank that consisted of pathologist pre-diagnosed ("gold-standard"), paraffin-embedded, human skin and lymph node tissues. Thin sections from each block were mounted on infrared reflective microscope slides and imaged, and a selection of the total images nominated as either training or test samples. Each training sample image was then compared to its corresponding haematoxylin and eosin (H&E)-stained section and reference library spectra extracted. Vertex component analysis (VCA) as a spectral feature extraction method was also explored. Classification of the test sample images was then performed using the spectral angle mapper (SAM) algorithm and the accuracy assessed by comparing the resulting classification images to the H&E-stained tissue sections. The tissue classification model developed produced a range in result quality, and highlighted various critical aspects in the construction of such methodologies. The taking of spectral derivatives improved image classification, as did the removal of paraffin from the tissue (although no data treatment targeting the paraffin was conducted on the non-deparaffinised tissues). Although the accuracy achieved in this study fell short of that required for clinical practice, the results obtained demonstrate that further investigation into the SAM algorithm as a tissue classifying tool is certainly warranted.

The second pathology application explored the ageing of wounds, a determination that may be critical in criminal investigations, particularly in homicide investigations, in which the timing of wound infliction may be crucial evidence. For this study, incised wounds were

inflicted on rats in a controlled manner and the tissue excised following a known amount of healing time ranging from 5 minutes to 288 hours (12 days). Thin sections of the wounds were mounted on infrared reflective slides, deparaffinised and then imaged using FTIR CI. Although four classification models were attempted, none were capable of producing highly accurate wound age determination. Spectral variation was observed between earlier and later wound ages using some of the classification methods, but the ability to correctly group the test samples into their respective age groups was not achieved. Based on the number of variables which must be taken into consideration when performing such a study, and the number of areas identified as needing further improvement (e.g. spectral data quality), the fact that even a limited form of discrimination was achieved using FTIR CI was encouraging.